# APPENDIX J

## OZONE MORTALITY META-ANALYSIS

# APPENDIX J.1
# ASSESSMENT AND SYNTHESIS OF AVAILABLE EPIDEMIOLOGICAL EVIDENCE OF MORTALITY ASSOCIATED WITH AMBIENT OZONE FROM DAILY TIME-SERIES ANALYSES

## 1.0 OVERVIEW OF AVAILABLE LITERATURE

### 1.1 Purpose and scope of this literature review

This document reviews and summarizes the available epidemiologic evidence concerning the relationship between ambient ozone concentrations and human mortality risks. The evidence is reviewed for the purposes of developing a quantitative procedure for estimating the change in number of premature deaths expected for each of the proposed NAAQS revision options. This quantification procedure is intended to be used as part of the Regulatory Impact Assessment (RIA) of the proposed NAAQS revisions. Its purpose, therefore, is to quantify, to the extent feasible given available information, the expected benefits of the proposed NAAQS revisions in terms of expected reductions in premature mortality throughout the country.

The literature relevant to this issue has been evolving rapidly, with many new research findings becoming available in the past two years. Many recent studies, therefore, are not discussed in the most recent version of the U.S. Environmental Protection Agency's criteria document for ozone (U.S. EPA, 1996). In order to take advantage of all available information, this review includes, but is not limited to, those studies discussed in the criteria document. The goals of the RIA process are to be as inclusive as possible in quantifying the expected costs and benefits of the proposed regulatory changes.

The goals of the RIA process may lead to a somewhat different approach to and interpretation of the literature than that taken in the criteria document and staff paper process. The focus of the latter is to determine what standard to set to protect public health, which is a somewhat different question than the quantification of the health effect changes expected as a result of one standard versus another. Given the goals of the RIA, the analysis errs on the side of reflecting all the expected benefits and costs, even if these cannot be precisely quantified.

This section reviews the literature on ozone and mortality and discusses key factors in evaluating the studies. This is followed by a discussion of the criteria used to select studies for a quantitative analysis of the association between ozone and mortality, and a review of each of the studies selected. The last section describes the quantitative method used to synthesize the information from multiple studies, including the methodology used to incorporate uncertainty into the analysis.

## 1.2 Overview of the literature on ozone and daily mortality

Table 1 lists 28 daily time-series epidemiology studies identified in the literature review that report results on a possible association between daily ozone concentrations and daily mortality. These studies were conducted in various urban areas throughout the world. Of these studies, 16 were conducted in the United States or Canada, 8 were conducted in Europe, and the remainder were conducted elsewhere, including Mexico City, Sao Paulo, Santiago, and Brisbane.

Of the 28 studies listed in Table 1, 21 were published or presented since 1995, illustrating the rapid expansion in this body of research in the last two and a half years. The studies show mixed findings as to whether there is a statistically significant association between daily ozone concentrations and daily mortality in each of the study areas. Overall, 15 of the studies report a statistically significant relationship between ozone and mortality, with the more recent studies tending to find statistical significance more often than the earlier studies.

## 1.3 Issues in evaluating the literature

Table 1 is a comprehensive list of studies, including some that report only qualitative results such as statements that ozone was not statistically significant, as well as conference papers and other manuscripts that may not be considered peer reviewed. As part of the process of evaluating the evidence presented by this body of literature, each study must be evaluated as to the soundness of the data and the analysis techniques and the conclusions drawn by the authors. This section highlights the key issues considered in the literature review. Section 2 goes into more detail as to the specific selection criteria developed for choosing studies to be used in the quantitative assessment.

Evaluation and interpretation of the studies is needed to assess the likelihood that the relationship between ozone and mortality is real. Studies must be assessed individually and as a whole to draw appropriate conclusions. Some of the key considerations for reviewing the studies are discussed in Bradford Hill's presidential address (Hill, 1965). Although he presents a number of guidelines for evaluating whether the associations observed in epidemiological studies are causal, he also notes that it is not necessary to meet or evaluate all of the guidelines before resolving to address the potentially harmful exposure. This review evaluates the studies in light of those principles most appropriate to the ozone mortality literature, but does not give a formal one-by-one evaluation of Hill's criteria.

The discussion below focuses on three key factors in evaluating the studies: whether covariates that might be confounding the ozone mortality relationship have been adequately addressed, whether the studies have sufficient data and statistical power to draw meaningful inferences, and whether there is consistency in findings across different studies as well as coherence between the findings in this body of literature and other research findings with regard to health effects of ozone.

**Table 1: Epidemiological Evidence on the Relationship Between Daily Mortality and Exposure to Ambient Ozone**

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Anderson et al., 1996 (APHEA project) | London 1987-1992 | 8-hr avg and daily 1-hr max (1-day lag) | 8-hr avg; with black smoke in the model: 1.029 (1.015 - 1.042) | included | Authors note that ozone effects remain signif. after $NO_2$ and $SO_2$ are added to the model. |
| Cifuentes and Lave, 1997 | Philadelphia 1983-1988 | daily 1-hr max | 1.008 (1.000 - 1.017) | excluded: not in or accepted to a peer reviewed journal | |
| Dockery et al., 1992 | St. Louis; Kingston- Harriman, TN Sept 1985-Aug 1986 | daily avg | St. Louis 1.0073 (0.9705 - 1.0735) East. TN 0.9839 (0.9017 - 1.0735) | excluded: no copollutants in model | |
| Hoek et al., 1997 (in press) | Rotterdam, The Netherlands 1986-1991 | daily avg | 1.044 (1.007 - 1.079) | included | Ozone was measured in $\mu g/m^3$. To convert to ppb, concentrations in $\mu g/m^3$ were divided by 1.96 (see note at end of table). |
| Ito and Thurston, 1996 | Cook County, Illinois 1985-1990 | 2-day avg | 1.017 (1.002 - 1.029) | included | |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Katsouyanni et al., 1993 | Athens, Greece<br>July 1987 | daily 1-hr max | (See comment) | excluded: no copollutants in model; July only | Ozone was measured in µg/$m^3$. To convert to ppb requires knowing the conversion factor for the temperature in the study, which was substantially higher (July only) than the other studies for which conversions were made.<br><br>The authors use smoke, $SO_2$, and ozone as alternative indices of air pollution but do not include more than one pollutant in any model. In each model, the air pollution index was binary. None of the air pollution indices was significant. |
| Kinney and Ozkaynak, 1991 | Los Angeles County<br>1970-1979 | daily 1-hr max (total oxidants) 1-day lag | 1.0059<br>(1.0033 - 1.0086) | excluded: measured oxidants, not ozone | This study measured total oxidants, rather than ozone specifically. |
| Kinney and Ozkaynak, 1992 (Abstr.) | New York City<br>April - Sept.<br>1971 - 1976 | daily 1-hr max | 1.0085<br>n.a. | excluded: not in or accepted to a peer reviewed journal | Although the authors do not report a standard error, they report that the coefficient was statistically significant at $p < 0.001$. |
| Kinney et al., 1995 | Los Angeles County<br>1985-1990 | daily 1-hr max | 1.000<br>(0.989 - 1.010) | included | |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Loomis et al., 1996 (HEI) | Mexico City 1991-1992 | daily 1-hr max | 0.995 (0.987 - 1.004) | included | This study presents results from many models; some results are based on daily 1-hr max ozone whereas others are based on daily avg ozone. The result shown here is from a model with both TSP and $SO_2$, using daily 1-hr max ozone. Results from this paper were, in general, not significant. |
| Moolgavkar et al., 1995 | Philadelphia 1973-1988 | daily avg, 1-day lag | 1.0154 (1.0045 - 1.0260) | included | |
| Ostro, 1995 | Southern California summers, 1980-1986 | daily avg | No PM proxy: 1.005 (1.000 - 1.012)<br>with estimated $PM_{2.5}$: 1.0025 (0.9951 - 1.010) | excluded: summer only | |
| Ostro et al., 1996 | Santiago, Chile 1989-1991 | daily 1-hr max | OLS: 0.986 (0.977 - 1.000)<br>Poisson regr : 0.995 (0.986 -1.005) | included | |
| Ozkaynak et al., 1995 (conf. paper) | Toronto, Canada 1972-1990 | daily 1-hr max | 1.0107 (1.0021 - 1.0194) | excluded: not in or accepted to a peer reviewed journal | |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Saldiva et al., 1994 | Sao Paolo, Brazil<br>May 1990-April 1991 | 3-day moving avg | 1.0315<br>(0.8933 - 1.1911) | excluded: not all population (age 5 or under) | |
| Saldiva et al., 1995 | Sao Paolo, Brazil<br>May 1990-April 1991 | daily avg and daily 1-hr max | daily avg: 0.9673<br>(0.8963 - 1.0438)<br><br>daily 1-hr max: 1.0099<br>(0.9896- 1.0307) | excluded: not all pop. (age 65+) | |
| Samet et al., 1996, 1997 (HEI) | Philadelphia<br>1974-1988 | 2-day avg | 1.024<br>(1.008 - 1.039) | included | This study considered ozone with only TSP and ozone with several other pollutants, including TSP (shown here). In both cases ozone was statistically significant. |
| Sartor et al., 1995 | Belgium<br>summer 1994 | daily avg | n.a. | excluded: does not report quantitative result; summer only | $O_3$ and temperature, both lagged one day, were correlated with daily mortality. This study focused primarily on a heat wave in the summer of 1994. Because temp. and ozone were highly correlated, "additive regression models with these two variables were unstable and unreliable, impeding to establish the true relationship between the number of daily deaths and these two environmental factors." |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Schwartz, 1991 | Detroit 1973-1982 | avg of 1-hr peaks; avg of daily means | n.a. | excluded: no co-pollutants in model; does not report quantitative result | Ozone was "highly insignificant as a predictor of daily mortality."<br><br>Unclear whether TSP and SO2 were also in the model containing ozone. |
| Shumway et al., 1988 | Los Angeles County 1970-1979 | avg of daily maxima at 6 monitors | n.a. | excluded: does not report quantitative results | Ozone was not included among the variables ultimately chosen for the regression models. The authors note the "near collinearity of temperature and ozone levels." |
| Simpson et al., 1997 | Brisbane, Australia 1987-1993 | ? | n.a. | excluded: not in or accepted to a peer reviewed journal | Described in Thurston, 1997: "When all pollutants [$SO_2$, $NO_2$, and particulate matter, indicated by nephelometer readings] were included in the model, only ozone and particulate matter remained significant." |
| Sunyer et al., 1996 (APHEA project) | Barcelona, Spain 1985-1991 | daily 1-hr max | 1.023<br>(1.0006 - 1.041) | excluded: no copollutants in model | Ozone was measured in $\mu g/m^3$. To convert to ppb, concentrations in $\mu g/m^3$ were divided by 1.96 (see note at end of table). Although several pollutants are considered, there do not appear to be any copollutant models; each pollutant seems to be considered in a separate model. |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Thurston 1997 (AWMA presentation) | Nine U.S. cities 1981-1990 | daily 1-hr max | Atlanta — 1.019 (1.012 - 1.027)<br>Chicago — 1.017 (1.012 - 1.022)<br>Detroit — 1.024 (1.017 - 1.032)<br>Houston — 1.005 (1.000 - 1.010)<br>Los Angeles — 1.007 (1.006 - 1.009)<br>Minneapolis — 1.017 (1.005 - 1.029)<br>New York — 1.019 (1.016 - 1.022)<br>San Francisco — 1.022 (1.008 -1.035)<br>St. Louis — 1.012 (1.006 - 1.019) | excluded: no copollutants in model; not in or accepted to a peer reviewed journal | 95% confidence intervals are calculated from reported relative risks and t statistics for the ozone coefficients in the log-linear regressions. |
| Touloumi et al., 1997 (in press) | Several European cities | daily 1-hr max | Fixed effects model: 1.018 (1.009 - 1.026)<br>Random effects model: 1.027 (1.005 - 1.049) | excluded: a meta-analysis of several cities; fixed effects assumptions rejected | This is a meta-analysis of the results from several European cities in the APHEA project. Random effects model found preferable for meta-analysis. Ozone was measured in µg/$m^3$. To convert to ppb, concentrations in µg/$m^3$ were divided by 1.96 (see note at end of table). |
| Verhoeff et al., 1996 | Amsterdam 1986-1992 | daily 1-hr max (2-day lag) | with black smoke: 1.014 (0.984 - 1.046)<br>with $PM_{10}$: 1.024 (0.974 -1.078) | included | Ozone was measured in µg/$m^3$. To convert to ppb, concentrations in µg/$m^3$ were divided by 1.96 (see note at end of table). |

| Study | Study Location/ Duration | $O_3$ Exposure Measure (ppb) | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$* | Included or Excluded (If Excluded, Reason) | Comments |
|---|---|---|---|---|---|
| Wyzga and Lipfert, 1995 | Philadelphia 1973-1990 | daily avg (1-day, 2-day, 3-day) | 1.0185 (n.a.) — 0 day lag<br><br>1.0308 (n.a.) — lag up to 1 day<br><br>1.0526 (n.a.) — lag up to 2 days | excluded: not in or accepted to a peer reviewed journal | Although standard errors corresponding to each reported coefficient were not given, a "typical" standard error of 0.0120 was given. The authors note that "the standard errors shown are reasonably constant over the range of lags." Using this std. error, all three relative risks are statistically significant. |
| Wyzga and Lipfert, 1996 | Philadelphia 1973-1980 | daily avg | age < 65:<br>1.024 (1.0003 - 1.0489)<br><br>age 65+:<br>1.0001 (0.9810 - 1.0197) | excluded: not in or accepted to a peer reviewed journal | This study considered only two separate age groups: < 65 and 65+, but not all pop. |
| Zmirou et al., 1996 (APHEA project) | Lyon, France 1985-90 | daily mean and daily 1-hr max | daily mean :<br>1.029 (0.951 - 1.117)<br>daily 1-hr max:<br>1.039 (0.941 - 1.157) | excluded: no copollutants in model | Ozone was measured in µg/$m^3$. To convert to ppb, concentrations in µg/$m^3$ were divided by 1.96 (see note at end of table). |

* Results are considered statistically significant if the lower bound of the 95% confidence interval is greater than or equal to one. Ozone was measured in $\mu g/m^3$ in the following studies: Verhoeff et al.,1996; Hoek et al., 1997; Sunyer et al., 1996; Touloumi et al., 1997; Katsouyanni et al., 1993; and Zmirou et al., 1996. To convert to ppb (vol.), concentrations in $\mu g/m^3$ were divided by 1.96 for all the studies except Katsouyanni et al., 1993. The conversion factor depends on the temperature in the study area. At 0˚ C (32˚ F) the factor is 2.144. Schwartz, 1995 (a study of respiratory hospital admissions in New Haven, Connecticut and Tacoma, Washington) used 1.96 to convert from $\mu g/m^3$ to ppb. The mean temperature in each of these cities was given as 52˚ F (about 11˚ C). The mean (or median) temperatures given in the other locations for which conversions were necessary were all about the same as in New Haven and Tacoma, except for in the Katsouyanni study, which considered only July [(Verhoeff et al.,1996 — 10˚ C; Hoek et al., 1997 — 10˚ C; Zmirou et al., 1996 — 12˚ C; Sunyer et al., 1996 — 11˚ C in winter and 20˚ C in summer; and Touloumi et al., 1997 — 13.3˚ C (an average of the six city-specific means given in the paper.)].

### 1.3.1 Covariates and other model specification issues

The accuracy of an estimate of a concentration-response function reported by a study depends on the study design. In general, critical considerations in evaluating the design of a daily time-series epidemiological study include the adequacy of the measurement of ambient ozone and the consideration of potentially important health determinants and confounding factors such as the following:

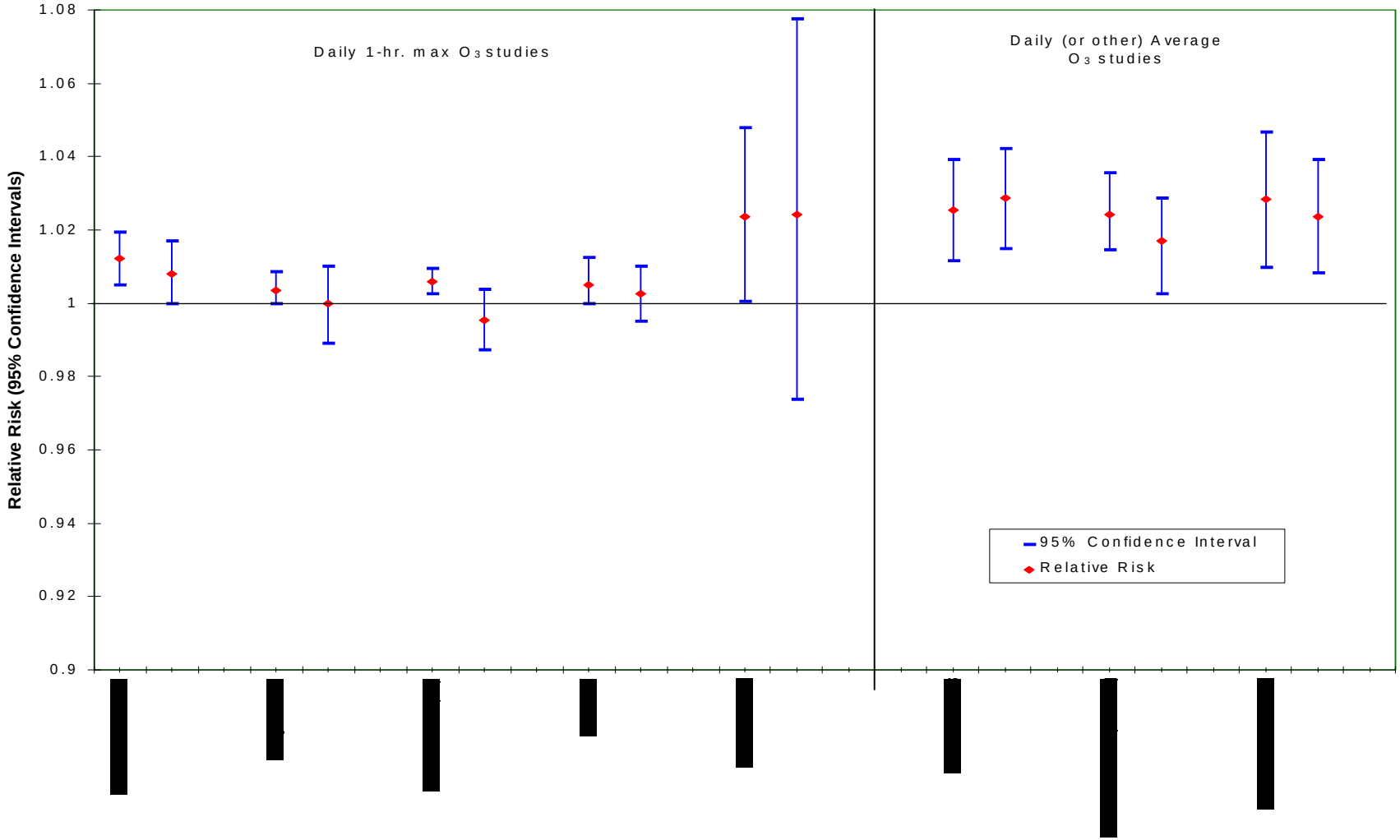
- copollutant air quality
- confounding effects of weather and season on mortality.

Ozone is a photochemically formed pollutant, so its presence and level of concentration are inevitably related to meteorological conditions, such as heat waves, that may themselves contribute to mortality risks, as well as to other air pollutants, such as some types of fine particulates, that tend to form and accumulate under similar conditions. Because particulate matter has been found to be associated with mortality risks, it is important to control for particulate matter concentrations when analyzing the potential mortality risk associated with ozone if the two pollutants are correlated. Some studies (listed in Table 1) have done this and some have not. Ozone and particulate matter are more highly correlated in some locations than in others. Examining results obtained for both pollutants in copollutant models across different locations may help sort out whether there appears to be an independent effect of each pollutant.

Figure 1 shows relative risks associated with a 25 ppb increase in ozone based on the results of single pollutant (ozone only) models and copollutant (ozone and PM or some proxy for PM) models from the same study, for several studies. In six out of the eight studies, the relative risk for ozone from a single pollutant model is higher than the relative risk for ozone from a copollutant model. In two cases (Anderson et al., 1996, and Verhoeff et al., 1996), however, the relative risk for ozone from the copollutant model is higher. The addition of PM or a proxy for PM to the model also increases the width of the 95% confidence interval of the relative risk for ozone in many (although not all) cases. The graph indicates that the association between ozone and mortality cannot be wholly attributable to confounding effects from PM.

All the studies reported in Table 1 have at least included a measure of daily temperature in the analysis, but the treatment of daily weather conditions and seasonal patterns in the data varies in the level of statistical sophistication. More recent studies tend to explore these potentially confounding factors more thoroughly with techniques such as smoothing to remove seasonal trends and nonlinear modeling to account for the effects of temperature extremes on daily mortality risks. A few authors have noted that a high level of collinearity between daily ozone and daily temperature in their data has made it difficult to draw conclusions about whether there may be an independent effect of ozone on mortality risk. For example, Sartor et al. (1995) noted that the high correlation between ozone and temperature made the models unstable when

**Figure 1: Relative Risks of Mortality Associated With a 25 ppb Increase in Ozone from Ozone-Only Models (Left Bar) and Copollutant Models (Right Bar)**



both of these variables were included, although the association was evaluated for only one summer. Shumway et al. (1988) also noted problems with high collinearity between temperature and ozone. All of the studies selected for the quantitative assessment included variables for daily temperature and seasonal trends in the ozone model. Sine/cosine functions have been frequently used to adjust for seasonal trends in daily mortality. Several of the studies, especially those using at least six years of daily data, found a statistically significant ozone effect after carefully controlling for daily weather conditions and seasonal patterns in the data.

Several of the studies considered different model specifications of weather, with varying degrees of complexity. The study that focused most particularly on the sensitivity of the air pollution-mortality relationship to different methods of controlling for weather (Samet et al., 1996, 1997) found that the approach used to characterize weather had no meaningful effect on relative risks associated with air pollution. This is consistent with a similar finding that the PM/mortality relationship was relatively insensitive to different methods of adjusting for weather (U.S. EPA, 1996).

### 1.3.2 Power of the studies

The effects of air pollutants on daily mortality incidence in a population are likely to be small compared to all the other factors that affect daily mortality incidence. It therefore takes a considerable amount of data for the effect to be measurable at a level of statistical significance, even if the effect is real. There is no set rule as to how many observations are needed, but most daily time-series studies use at least one to three years of data. The datasets have tended to become larger in more recent studies as longer series of air quality monitoring data have become available over time.

Table 2 illustrates an interesting relationship across the studies in Table 1; those with more years of data tend to be more likely to report a statistically significant ozone effect on daily mortality. Of the seven studies since 1995 that report finding no statistically significant ozone effect, the average number of years of data was less than four. Of the 13 studies since 1995 that report a statistically significant ozone effect, the average number of years of data was about 10. A similar pattern is seen in the pre-1995 studies. Table 2 is a simplistic analysis of this issue, but the pattern illustrated is suggestive of the possibility that it takes many years of data before the ozone effect can be separated from the daily weather and seasonal patterns with which it tends to be correlated.

### 1.3.3 Consistency and coherence of the studies

Consistent and coherent epidemiological findings in repeated studies in various locations also support an inference of causality in the observed relationship between ozone and mortality. Thus, a key consideration in the evaluation of the epidemiological studies is the consistency and coherence of the findings to date. The evaluation of consistency and coherence relies on a series of judgments of the quality literature and thus must be a qualitative one.

Consistency refers to whether the estimated relationship between ozone and mortality is similar across different locations and circumstances. This does not mean that the findings have to be identical for each study because there are many legitimate reasons why an ozone effect on mortality incidence may vary from location to location, including variations in lifestyle and climate that may affect the population's exposure to outdoor air pollutants. However, if there truly is a causal relationship, we would expect to see repeated findings of a statistically significant relationship of a reasonably comparable magnitude in many different studies.

**Table 2. Years of Data Used versus Finding a Statistically Significant Ozone Coefficient**

| Study Finding Regarding Ozone Effect | Studies Released Before 1995 | | Studies Released 1995 to 1997 | |
|---|---|---|---|---|
| | Number of Studies* | Mean Number of Years of Data** | Number of Studies | Mean Number of Years of Data |
| No Statistically Significant Effect | 4 studies | 3.0 years | 7 studies | 3.7 years |
| Statistically Significant Effect | 2 studies | 8.0 years | 13 studies | 9.9 years |

*Two studies did not report quantitative results or did not report sufficient information to determine whether results were statistically significant. The total number of studies in this table therefore is 26.

** A few studies reported that their data sets contained less data than the number of calendar years that they covered because of missing data. For example, Kinney et al. (1995) used six years of calendar data, but had only every six day $PM_{10}$ measurements; their effective amount of data was therefore only one year. The figures reported here reflect the effective amount of data when sufficient information is reported by the authors to make an adjustment.

Of those studies selected (discussed in the following section) that meet the criteria for the quantitative assessment, several do not find a statistically significant effect of ozone on mortality. Of those that find a statistically significant effect, the magnitude of the effect is roughly comparable. Figures 2 and 3 show the estimated relative risk and 95% confidence interval for a 25 ppb change in ozone concentration from each of the studies in Table 1 that reports quantitative results. These figures show both the results selected for the quantitative assessment as well as those that were eliminated for one reason or another. Figure 2 shows the results estimated for daily high-hour ozone concentrations. Figure 3 shows the results estimated for a daily average ozone concentration. These are shown on separate graphs because a 25 ppb change in the daily high-hour is not comparable to a 25 ppb change in the daily average; these two sets of relative risk results, therefore, are not directly comparable.

The studies are suggestive of a consistent association, but uncertainties remain. It is possible that the inconsistencies between the studies are due to some of the factors discussed

**Figure 2: Results of Epidemiology Studies on Daily 1-Hour Maximum Ozone and Daily Mortality: Relative Risk and 95% CI for a 25 ppb Increase in $O_3$**

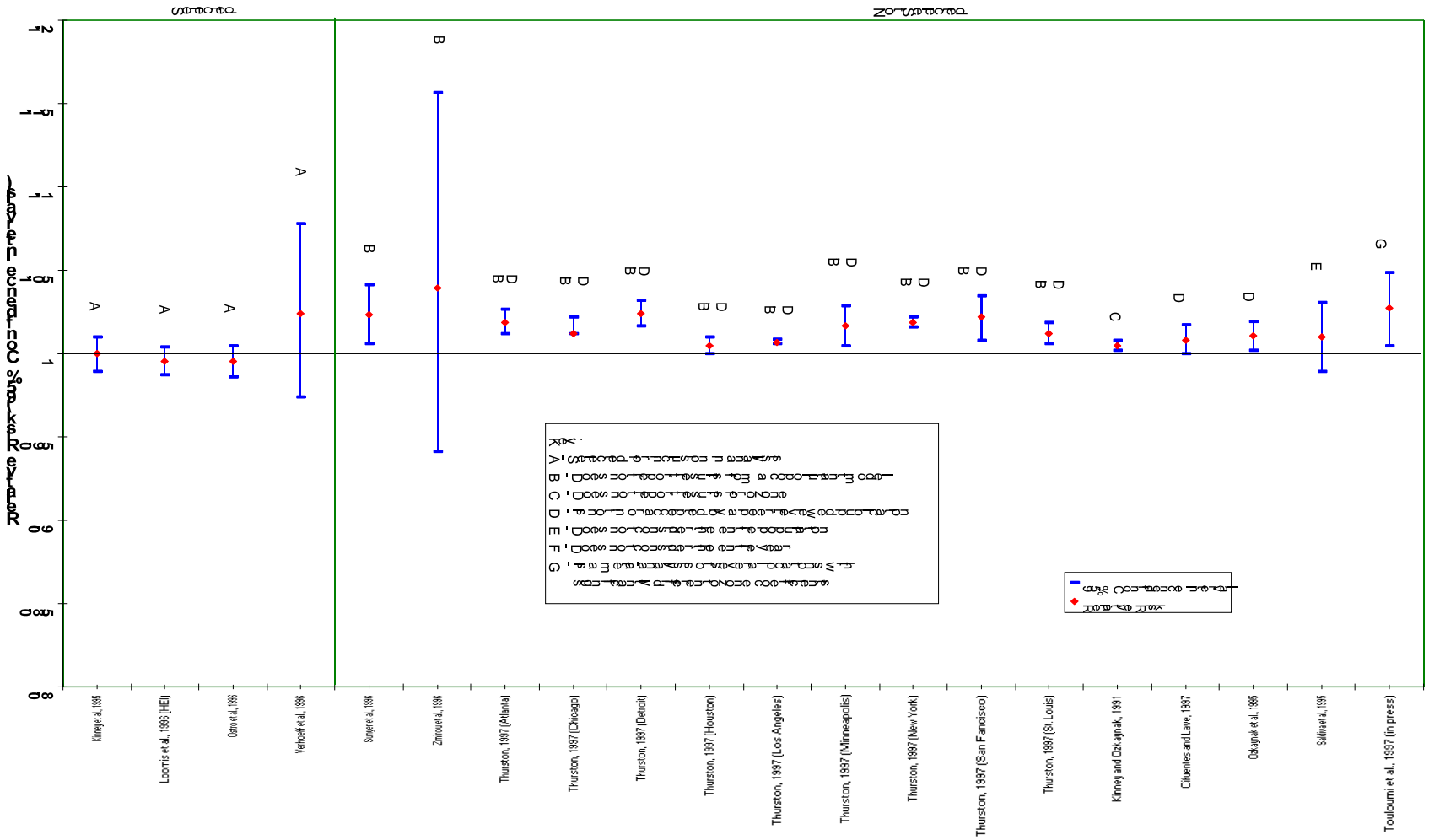
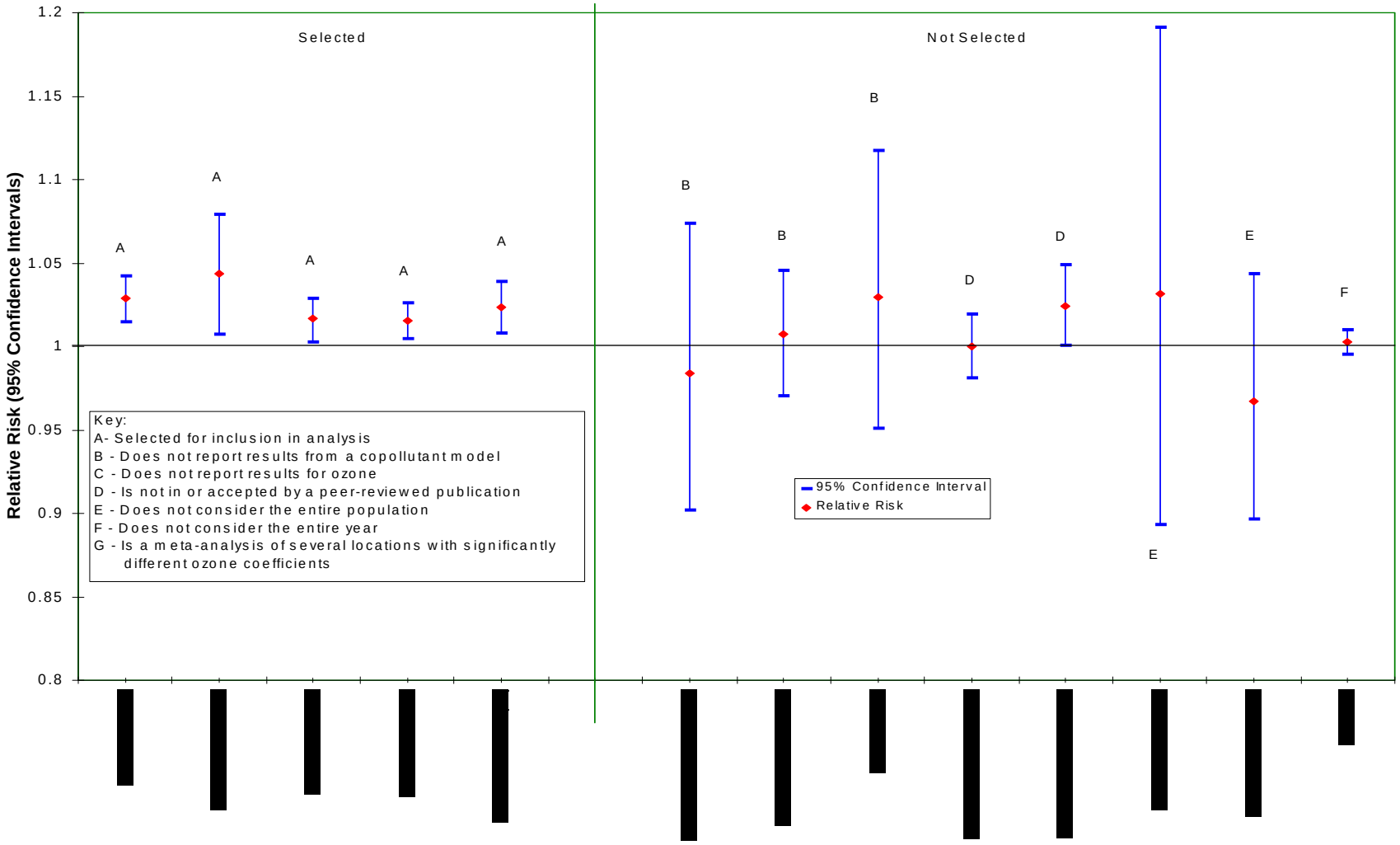


**Figure 3: Results of Epidemiology Studies on Daily (or Other) Average Ozone and Daily Mortality: Relative Risk and 95% CI for a 25 ppb Increase in $O_3$**



previously, such as statistical power. Nevertheless, although the studies point to the likelihood of an effect, it still cannot be concluded that there is an unambiguously statistically significant relationship between ozone and mortality. The quantitative assessment methodology takes the uncertainties into account by reflecting the full range of the findings, including those studies that do not find a statistically significant ozone-mortality relationship.

Coherence refers to whether the ozone mortality relationship makes sense given other available evidence concerning health effects associated with ozone exposure. All of this evidence is fully reviewed in the ozone criteria document (U.S. EPA, 1996). There is a wide range of evidence that ozone is an irritant to the respiratory system, including evidence from clinical, laboratory, and epidemiologic studies. Health effects of a serious nature have been associated with daily fluctuations in ozone concentrations, including respiratory hospital admissions and aggravation of chronic respiratory diseases such as asthma. There is at this time no clearly delineated biological mechanism that can explain how ozone exposure may result in premature mortality, but it is plausible that such a relationship may exist given the other available evidence of ozone health effects.

### 1.4 Key uncertainties in ozone mortality benefits estimation

For a quantitative assessment of the reduction in premature mortality expected as a result of the proposed changes to the NAAQS, we want to answer the following question: How many additional premature deaths will be avoided for each of the standard alternatives under consideration, relative to what would be expected under current standards?

One of the strengths of epidemiology studies is that they analyze actual mortality incidence in human populations at ambient pollution concentrations. Subjects are studied in their normal environment and the mortality incidence is directly observed. A major challenge for epidemiology studies is the difficulty in isolating with confidence the effects of a specific air pollutant such as ozone when this may be just one of many complex factors that influence human mortality.

Any quantitative mortality risk assessment faces challenges in the form of incomplete information, making it necessary to employ a variety of assumptions. Most of the assumptions necessary in the quantitative assessment of changes in mortality incidence in association with alternative NAAQS for ozone are the same as those required for all of the health effects estimates based on epidemiologic studies.

The concentration-response function is a key element of the quantitative assessment. The accuracy the assessment depends, in part, on (1) how well the concentration-response functions used in the assessment have been estimated (e.g., whether they are unbiased estimates of the relationship between mortality and ambient ozone concentration in the original study locations), (2) how applicable these functions are to locations and times other than those in which they were estimated, and (3) the extent to which these relationships apply to the range of the ozone concentrations to which they are being applied in the assessment.

## 2.0 SELECTION CRITERIA FOR QUANTITATIVE ANALYSIS

Several criteria were used to select studies for inclusion in the quantitative analysis, and, within selected studies, to select from among several reported results. A study was included in the quantitative analysis only if it:

1. measures daily mortality (i.e., is a time series study)
2. reports quantitative results for ozone
3. is in or has been accepted by a peer-reviewed publication
4. reports results from a copollutant model, including PM or some proxy for PM in the model with ozone, as well as some measure of temperature and season
5. considers the entire population (rather than only a subset of the population) in the study location
6. considers the whole year (rather than only a season or seasons)
7. considers only a single location or, if it is a meta-analysis of several locations, has been unable to reject the hypothesis that the ozone coefficient is the same in all locations considered.

The reasonable selection of a single ozone result from among two or more ozone results reported in the same study is facilitated, in almost all cases, by the following two criteria:

8. PM ($PM_{10}$ or $PM_{2.5}$) is preferable to other measures of particulate matter
9. More pollutants in the model is preferable to fewer pollutants.

Each of these study and result selection criteria is discussed more fully below. Reporting a statistically significant positive result for ozone is not a criterion for study selection, nor does statistical significance or size of coefficient (or relative risk) affect the criteria for result selection within a study. Table 3 lists the final selection of studies and, for each study selected, the final selection of ozone results. Figure 4 depicts the selected ozone results as the relative risks (and 95% confidence intervals) associated with a 25 ppb increase in ozone. Figures 2 and 3 show that the studies eliminated because of the quantitative assessment selection criteria did not on the whole find substantially higher or lower ozone effects than the selected studies.

**Table 3: Studies and Ozone Results Selected for Quantitative Analysis of the Relationship between Daily Mortality and Exposure to Ambient Ozone***

| Study | Study Location/ Duration | Copollutants in model | Means and Ranges (or Std. Devs.) of PM (or Proxy) ($\mu g/m^3$) and $O_3$ (ppb)** | $O_3$ Exposure Measure (ppb)** | Ozone Lag | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$** | Comments |
|---|---|---|---|---|---|---|---|
| Anderson et al., 1996 (APHEA project) | London 1987-1992 | black smoke | black smoke: 14.6 ( 3-95)<br>$O_3$ 8-hr avg: 15.5 (1-74) | 8-hr avg (1 day lag) | none | 1.029 (1.015 — 1.042) | Authors note that ozone effect remains significant after $NO_2$ and $SO_2$ are added to the model. |
| Kinney et al., 1995*** | Los Angeles County 1985-1990 | $PM_{10}$ | $PM_{10}$: 58 (15-177)<br>$O_3$ 1-hr max: 70 (3-201) | daily 1-hr max | 1-day lag | 1.000 (0.989 — 1.010) | |
| Loomis et al., 1996 (HEI)**** | Mexico City 1991-1992 | TSP, $SO_2$ | TSP: n.a. (86-460)<br>$O_3$ 1-hr max: n.a. (26-319) | daily 1-hr max | none | 0.995 (0.987 — 1.004) | (Results are taken from Table 14 in the paper.) |
| Ostro et al., 1996**** | Santiago, Chile 1989-1991 | $PM_{10}$ | $PM_{10}$: 115 (30-367)<br>$O_3$ 1-hr max: 53 (11-264) | daily 1-hr max | 1-day lag | 0.995 (0.986 — 1.005) | The result from the Poisson regression model has been chosen. |

| Study | Study Location/ Duration | Copollutants in model | Means and Ranges (or Std. Devs.) of PM (or Proxy) ($\mu g/m^3$) and $O_3$ (ppb)** | $O_3$ Exposure Measure (ppb)** | Ozone Lag | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$** | Comments |
|---|---|---|---|---|---|---|---|
| Verhoeff et al., 1996**** | Amsterdam 1986-1992 | $PM_{10}$ | $PM_{10}$: 38 (n.a.-191)<br>$O_3$ 1-hr max: 22 (n.a. — 154) (converted from $\mu g/m^3$ to ppb) | daily 1-hr max | 2-day lag | 1.024 (0.974 -1.078) | $O_3$ was measured in $\mu g/m^3$. To convert to ppb, ozone concentrations in $\mu g/m^3$ were divided by 1.96 (see note at the end of the table).<br><br>The result from the model with $PM_{10}$ (rather than black smoke) has been chosen. |
| Hoek et al., 1997 (in press) | Rotterdam, The Netherlands 1986-1991 | TSP (1 day lag)<br><br>$SO_2$ (1 day lag) | TSP: 42 (21 — 287)<br>$O_3$: 13.7 (0.5 — 67.3) (converted from $\mu g/m^3$ to ppb) | daily avg | 1-day lag | 1.044 (1.007 — 1.079) | $O_3$ was measured in $\mu g/m^3$. To convert to ppb, ozone concentrations in $\mu g/m^3$ were divided by 1.96 (see note at the end of the table). |
| Ito and Thurston, 1996 | Cook County, Illinois 1985-1990 | $PM_{10}$ | $PM_{10}$: 40.7 ± 19.1<br>$O_3$: 38.1 ± 19.9 | 2-day avg | avg of 0-day and 1-day lags | 1.017 (1.002 — 1.029) | |

| Study | Study Location/ Duration | Copollutants in model | Means and Ranges (or Std. Devs.) of PM (or Proxy) ($\mu g/m^3$) and $O_3$ (ppb)** | $O_3$ Exposure Measure (ppb)** | Ozone Lag | Relative Risk and 95% CI for a 25 ppb Increase in $O_3$** | Comments |
|---|---|---|---|---|---|---|---|
| **The following studies will be used to generate a single distribution for Philadelphia:** | | | | | | | |
| Moolgavkar et al., 1995*** | Philadelphia 1973-1988 | TSP, $SO_2$ | TSP: n.a. (14.5 — 338)<br>$O_3$: n.a. (0.0 — 159) | daily avg | 1-day lag | 1.0154 (1.0045 — 1.0260) | |
| Samet et al., 1996, 1997 (HEI)**** | Philadelphia 1974-1988 | TSP, $SO_2$, $NO_2$, LCO | TSP: 67.3 (14.5 — 222.0)<br>$O_3$: 19.8 (0 — 90.0) | 2-day avg | avg of 0-day and 1-day lags | 1.024 (1.008 — 1.039) | The result from the model with several copollutants has been chosen. |

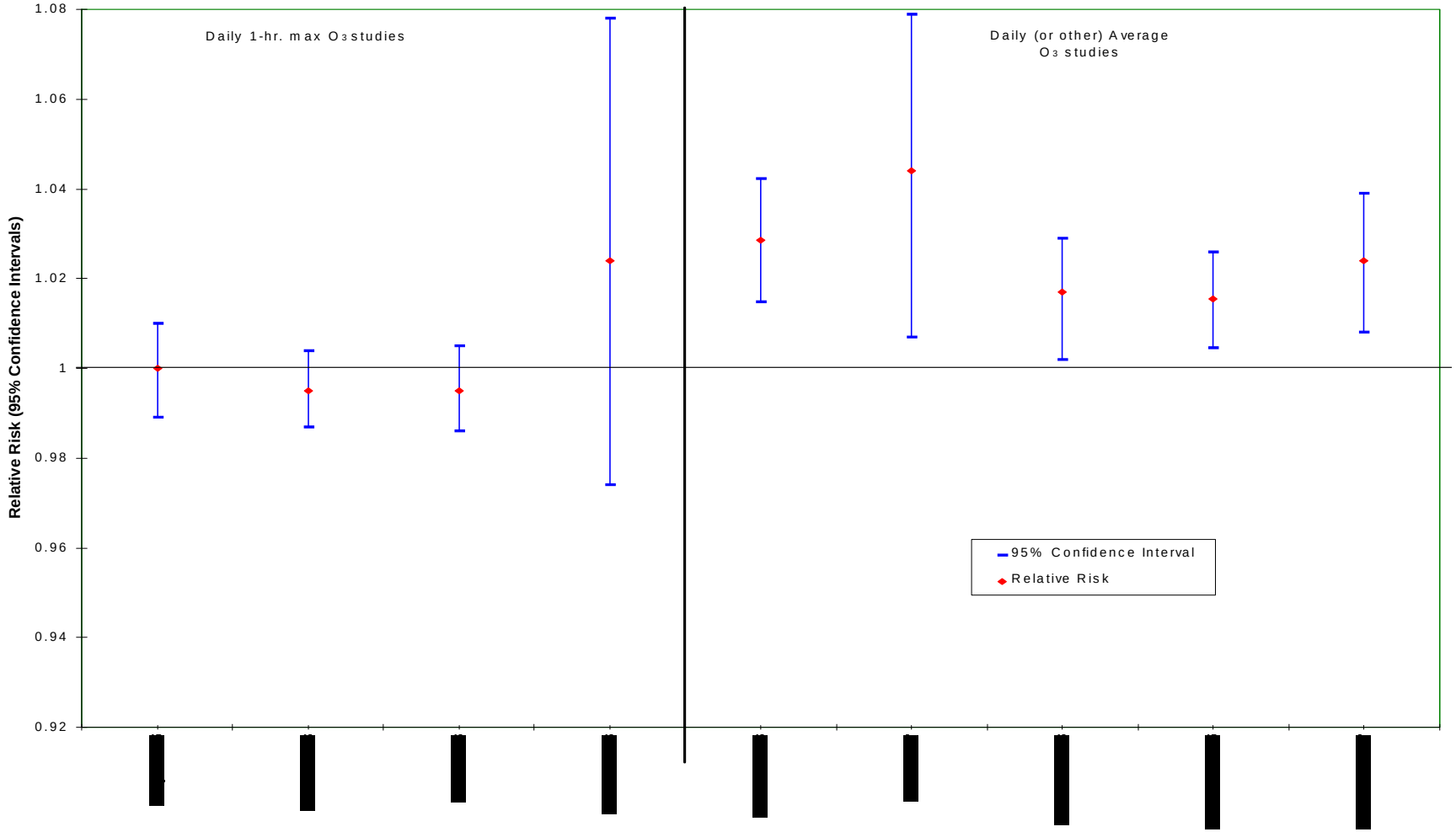
* To be included among the studies used in the quantitative analysis, a study must satisfy all the study selection criteria; one study (Samet et al., 1996) satisfied the study selection criteria but was omitted from the quantitative analysis to avoid redundancy problems.

**Ozone was measured in $\mu g/m^3$ in Verhoeff et al.,1996, and Hoek et al., 1997. To convert to ppb (vol.), concentrations in $\mu g/m^3$ were divided by 1.96. The conversion factor depends on the temperature in the study area. At 0˚ C (32˚ F) the factor is 2.144. Schwartz, 1995 (a study of respiratory hospital admissions in New Haven, Connecticut and Tacoma, Washington) used 1.96 to convert from $\mu g/m^3$ to ppb. The mean temperature in each of these cities was given as 52˚ F (about 11˚ C). The mean (or median) temperatures given in the other locations for which conversions were necessary were about the same as in New Haven and Tacoma (Verhoeff et al., 1996 — 10˚ C; and Hoek et al., 1997 — 10˚ C).

***Results from study were quantified and monetized in the December 1996 RIA in support of Ozone NAAQS.
Neither the 812 Retrospective study (October 1996, Draft) nor the Ozone Staff Paper Risk Assessment use or cite any ozone-mortality studies.

****This study presents more than one ozone result. A single result has been selected (see Section 2.5: Selecting from among multiple results reported from a study: Criteria 8 and 9).

**Figure 4: Results of Studies Selected for Inclusion in Quantitative Analysis: Relative Risks and 95% CI for a 25 ppb Increase in $O_3$**



## 2.1 Basic inclusion criteria: Criteria 1, 2, and 3

The first three criteria are the basic criteria used to ensure that the studies have usable (quantitative) results on daily mortality and that the analyses are of peer-reviewed quality. To require that the results be quantitative does not need comment. There are a few studies that estimate the relationship between long-term (e.g., annual) ozone and mortality using cross-sectional analyses. Although these may be valid studies, the majority of studies estimate the relationship between daily ozone and daily mortality. Because the relationship between long-term ozone and mortality may be different from that between daily ozone and mortality, these two types of studies cannot be combined in a quantitative analysis. The preponderance of daily studies therefore indicates the exclusion of the long-term studies from this analysis.

## 2.2 Consideration of key covariates: Criteria 4, 8, and 9

### 2.2.1 Copollutants

There is substantial evidence that there is a relationship between particulate matter (PM) air pollution and mortality. Although there is always a potential problem of separating the effects of confounded copollutants, the evidence for an association between PM and mortality is, to date, greater than for other pollutants in the typical "air pollutant mix." To the extent that PM and ozone are correlated, omitting PM from the model could tend to bias the estimate of the ozone coefficient. To avoid falsely attributing PM effects to ozone, it was required that any ozone result included in the analysis be from a model that included PM or some proxy for PM in the model (Criterion 4). (See Figure 1 for a comparison of ozone results with and without a PM measure in the model.) Because PM is clearly preferable to a proxy for PM, when there was a choice between the two, the model with PM was chosen (Criterion 8).

For the same reason that models with PM or a proxy for PM were required, models with more pollutants were preferred over models (in the same study) with fewer pollutants (Criterion 9). The omission of a pollutant that may be correlated with both ozone and mortality could result in biased estimates of the ozone coefficient (and other coefficients) in the model. Because the evidence for an association with mortality is strongest for PM, the inclusion of other pollutants in the model was not a study selection criterion; however, in choosing among multiple models within a single study, it provides a reasonable criterion by which to select a single, preferable result.

### 2.2.2 Weather and seasons

Many studies have noted the correlation between weather variables (temperature, in particular) and ozone, as well as the seasonality of ozone (which peaks in the summer). Because temperature is known to be associated with mortality, and because mortality is known to have an annual cycle, it is essential that temperature and seasonality be controlled in the model.

## 2.3 Consistency criteria: Criteria 5 and 6

Like the first criterion, Criteria 5 and 6 ensure that all the studies included in the quantitative analysis are measuring the same thing: the relationship between daily ozone concentrations and daily mortality for the entire population in a location over the course of the year. The few studies that analyzed separate seasons do not show a consistent pattern. In addition, statistically significant results are not limited to the "high" ozone season. Limiting the analysis to individual seasons, however, substantially reduces the number of observations, thereby also reducing the power of the analysis to detect effects. For consistency with the annual air quality analysis of proposed changes to the NAAQS, and to ensure that a minimum of a full year of data are included in the analysis, Criterion 6 requires that the result is based on a full-year analysis rather than a season-specific analysis. Because most studies that report season-specific results also report full-year results, this criterion serves more as a means of selecting from among multiple results from within a study than as a means of selecting studies.

Most studies estimated the ozone-mortality relationship for the entire population in a location. Because this relationship may vary by age group, it would be inconsistent to pool the results of all-population studies with those from studies limited to certain age groups. Therefore Criterion 5 requires that the results apply to the entire population rather than to a subset of the population.

### 2.4 Ensuring that each study represents a single location: Criterion 7

Criterion 7 excludes any study that is itself a pooling of the results from several separate locations with statistically significantly different ozone-mortality relationships. The reason for this criterion goes beyond consistency and has to do with the necessity of assigning weights to the studies that are input to the quantitative analysis, described in Section 4. There is no clear way to assign a weight to a meta-analysis when pooling its results with the results of single-location studies. (If the meta-analysis was unable to reject the hypothesis that all locations considered were the same, then it could be treated as a single-location study, making it roughly comparable to other single-location studies.)

In pooling the results of multiple studies, the quantitative analysis (described in Section 4) assigns a random effects weight to each study. Random effects weights are designed to give less weight to a coefficient from a study the greater its standard error, and to reflect the lower reliability of estimates with larger standard errors. The standard error associated with an estimate is the square root of the "within-study variance" of a single-location study. The greater the within-study variance, the less reliable the study result. The within-study variance of a meta-analysis, however, has two components: the standard errors of the estimates of location-specific coefficients in the meta-analysis (the within-study variances), and the between-location variance. Only the first component reflects true uncertainty and the corresponding lack of reliability of estimates. The second component reflects actual variability among locations. This variability among locations within the meta-analysis does not imply uncertainty or a lack of reliability. The "within-study variance" of a meta-analysis is therefore not comparable to the within-study variance of a single-location study. A random effects weight applied to a meta-analysis as if it

were a single-location study would therefore give too little weight to the study. The correct weighting scheme to incorporate the results of a meta-analysis is unclear.

Only one ozone-mortality meta-analysis (Touloumi et al., 1997) is in the literature. Because this study rejected the hypothesis that the ozone-mortality relationship is the same in all the locations considered (several cities in Europe), it cannot be included in the quantitative analysis as if it were a single location. The results of this study, however, are presented separately.

### 2.5 Selecting from among multiple results reported from a study: Criteria 8 and 9

Several studies report results from more than one analysis satisfying the first seven criteria (e.g., if they consider several different models). Loomis et al. (1996) report relative risks from many different models. However, according to Criterion 9, the result from a model that includes both TSP and $SO_2$ is selected over the other results reported in that study. Verhoeff et al. (1996) report an ozone relative risk from a model in which particulate matter is measured as black smoke and one from a model in which particulate matter is measured as $PM_{10}$. According to Criterion 8, the result from the model with $PM_{10}$ is selected. Samet et al. (1997) report an ozone result from a model including TSP and one from a model in which TSP, SO2, NO2, and LCO are all included in the model with ozone. According to Criterion 9, the latter result is selected.

After consideration of Criteria 8 and 9, only one study among those satisfying Criteria 1 through 7 with more than a single ozone result remains. Ostro et al. (1996) report results from both OLS and Poisson regression models. Because there is no obvious way to select between these two models, consistency with other studies is used as a criterion. Because the Poisson regression model is more commonly used, the result from the Poisson regression model is selected over the OLS result in Ostro et al. (1996). (Because the results of the two models are very similar, the selection of one model over the other will make little difference in the analysis.)

## 3.0 SUMMARY OF SELECTED STUDIES

### 3.1 Anderson et al., 1996. Air pollution and daily mortality in London: 1987-92

This study was one of several studies in the APHEA collaborative project investigating the relationship between daily levels of air pollutants and daily mortality in cities throughout Europe. A time series of daily counts of mortality for all ages and all causes of death except accidents was constructed for the Greater London area from April 1987 to March 1992. Data on mean daily temperature and humidity were obtained from central London. Both eight hour (9 a.m. to 5 p.m.) average and daily 1-hour maximum ozone concentrations were measured at a single monitor located near Victoria station in central London. Concentrations of nitrogen dioxide (daily average and daily 1-hour maximum), black smoke (daily average), and sulfur dioxide (daily average) were also monitored. The statistical analysis followed the APHEA protocol. Daily death count was the dependent variable in autoregressive log-linear regression

models with Poisson errors ("Poisson regression" models). Independent variables included temperature and humidity and the various pollution variables (although not all pollution variables were included in all models), as well as adjustments for time and seasonal trends, day of the week, holidays, and an influenza epidemic that occurred during the study period. The authors reported that the "U-shaped" relation between daily temperature and mortality was best adjusted for by fitting three separate linear terms for (1) < 5° C, (2) > 20° C, and (3) 5-20° C. Relative humidity was adjusted for by a single linear term.

Both 8-hour average and 1-hour maximum ozone were considered in single-pollutant models; 8-hour average ozone was also considered in models that included black smoke. In the single-pollutant models, both 8-hour average and daily 1-hour maximum ozone were statistically significant. The authors reported a relative risk of 1.024 [95% C.I. (1.011, 1.038)] when ozone was measured as an 8-hour average, and a relative risk of 1.026 [95% C.I. (1.013, 1.039)] when ozone was measured as a daily 1-hour maximum, associated with an increase from the 10th to the 90th percentile concentration (an increase of 24 ppb for the 8-hour average and 31 ppb for the 1-hour maximum). When black smoke was also in the model with 8-hour average ozone, the relative risk became 1.027 [95% C.I. (1.014, 1.041)].

### 3.2 Kinney et al., 1995. A sensitivity analysis of mortality/$PM_{10}$ associations in Los Angeles

This study investigated the relationship between daily mortality (excluding suicides, accidental deaths, and nonresident deaths) and daily pollution in Los Angeles County from January 1, 1985, to December 31, 1990. Because the focus of the paper was on the sensitivity of the $PM_{10}$/mortality associations to the analytic methods used, however, the time series of data used comprised only those days with $PM_{10}$ data (364 days). Death counts were obtained from the National Center for Health Statistics death certificate tapes. Data on 24-hour average $PM_{10}$ (collected every six days), daily 1-hour maximum $O_3$, and daily 1-hour maximum CO were obtained from the U.S. Environmental Protection Agency's Aerometric Information and Retrieval System (AIRS). The $PM_{10}$ data were taken from monitors at four sites, and the data on $O_3$ and CO were each collected from monitors at eight sites. In a single pollutant model with only ozone, and with same-day temperature and relative humidity and sine/cosine functions to adjust for seasonal trends, the relative risk associated with a 143 ppb increase in ozone was 1.02 [95% C.I. (1.00, 1.05)]. In a copollutant model with both ozone and $PM_{10}$, and with the same weather and season variables as in the single pollutant model, the relative risk became 1.00 [95% C.I. (0.94, 1.06)]. The authors concluded that "the $O_3$ effect on mortality, if any, is weaker than that of $PM_{10}$."

### 3.3 Loomis et al., 1996. Ozone exposure and daily mortality in Mexico City: A time-series analysis

This study was conducted under the auspices of the Health Effects Institute (HEI). Daily death counts in Mexico City's Federal District (which includes about half of the population of

Mexico City's metropolitan area) were obtained for the period from 1990 through 1992 from the Instituto Nacional de Estadistica, Geografia, e Informatica. Daily levels of $SO_2$, CO, $O_3$, and nitrogen oxides, as well as several meteorological variables, were taken from nine monitoring stations. TSP was measured at 19 monitoring stations. Although the authors describe the "basic metric of exposure" to ozone as the daily 1-hour maximum, four other measures of ozone, including the daily average, were also examined. The basic model used was the Poisson regression model. Daily death counts were regressed on pollution variables as well as minimum temperature and several time-related variables, including a sine-cosine function to remove seasonal trends. The results of a variety of models (considering different types of mortality, different age groups, different regions within Mexico City, different lag structures for pollutants, different combinations of pollutants and other variables, different measures of ozone, and variations on the functional form) are reported. Although the ozone effect was positive and statistically significant in a few single pollutant models, it was generally not significant in copollutant models.

### 3.4 Ostro et al., 1996. Air pollution and mortality: Results from a study of Santiago, Chile

This study investigated the relationship between daily death counts in metropolitan Santiago (excluding accidental deaths and deaths of residents that occurred outside the metropolitan area) and daily air pollution levels for 1989 through 1991. Because of missing data, however, data were available for all pollutants and for weather variables on a total of 779 days. Data were collected on $PM_{10}$ (daily average), $SO_2$ (1-hour maximum), $NO_2$ (1-hour maximum), and $O_3$ (1-hour maximum), as well as daily minimum and maximum temperature and daily average humidity. Ordinary Least Squares (OLS) regression and Poisson regression were both used to examine the relationship between mortality and air pollution. Although most emphasis was placed on $PM_{10}$, the investigators conducted several sets of sensitivity analyses, one of which considered the effects of pollutants other than $PM_{10}$. Each of the other pollutants was considered both with and without $PM_{10}$ in the model. In all cases, one-day lagged minimum temperature and binary variables for the hottest and coldest 10% of the days, as well as seasonal adjustments, were included. The relative risk associated with an increase of 52.8 ppb ozone in the single pollutant (ozone only) model for the summer only was 1.02 and was marginally statistically significant [95% C.I. = (1.00-1.05)]. The relative risk from the corresponding copollutant model (summer only) was still 1.02 but was no longer statistically significant. Relative risks from full-year models were not significant.

### 3.5 Verhoeff et al., 1996. Air pollution and daily mortality in Amsterdam

Daily death counts for the city of Amsterdam were obtained from the Municipal Population Register for the period 1986-1992. (Because the mortality data did not contain cause of death, accidental deaths presumably were not removed from the data.) Air pollution monitoring data were obtained from the Amsterdam Environmental Research Institute. Black smoke was measured daily at four sites throughout the study period; TSP was measured every 3 days at four sites from 1986 to 1988; in 1988 this was changed to $PM_{10}$. $SO_2$, CO, and $O_3$ were

measured continuously at 11, 5, and 5 sites, respectively. Because TSP and $PM_{10}$ were highly correlated during the period in which both were measured (Pearson correlation coefficient = 0.95), TSP concentrations measured during 1986-1988 were converted into $PM_{10}$ concentrations via linear regression. With the exception of $O_3$, which was measured as the daily 1-hour maximum, all pollutants were measured as daily averages. Daily average concentrations of black smoke and TSP were also available in the city of Rotterdam (about 80 km from Amsterdam) during the entire study period. Because these were correlated with the corresponding concentrations in Amsterdam (r = 0.6 for black smoke and r = 0.85 for TSP), the Rotterdam data were used to predict daily concentrations of black smoke and $PM_{10}$ in Amsterdam when local data were unavailable.

The effects of air pollution on daily mortality counts were examined using Poisson regression in models that included, in addition to pollutants, both weather variables and variables to account for seasonal and other time trends. Two dummy variables were created to characterize "warm" and "cold" days. The "warm" dummy was set to 0 if temperature was $\leq$ 16.5° C; it was set to temperature minus 16.5° C otherwise. The "cold" dummy was similarly assigned a zero if temperature was $\geq$ 16.5° C and was set equal to 16.5° C minus temperature otherwise. Temperature lags of up to 2 days were considered. When same-day ozone, 1 day lagged ozone, and 2 day lagged ozone were each considered in single pollutant models, the relative risks associated with a 100 $\mu g/m^3$ increase in ozone were all greater than 1.0 (1.018, 1.001, and 1.049, respectively). However, only the relative risk for the 2 day lagged ozone was statistically significant. When 2 day lagged ozone and $PM_{10}$ were both included in a model, the relative risk associated with a 100 $\mu g/m^3$ increase in ozone was 1.050, although it was not statistically significant.

### 3.6 Hoek et al., 1997 (in press). Effects of ambient particulate matter and ozone on daily mortality in Rotterdam, the Netherlands

This study investigated the relationship between daily mortality and daily air pollution in Rotterdam from 1983 to 1991, although consistent data on the gaseous pollutants ($SO_2$, CO, and $O_3$) were available from only 1986 to 1991. Daily death counts, including only deaths of residents of Rotterdam but without information on cause of death, were obtained from the Municipal Registry of the city of Rotterdam. Data on daily TSP and black smoke concentrations were obtained from the Rijnmond Environmental Protection Agency (DCMR). Data on daily $SO_2$, CO, and $O_3$ were obtained from the National Air Quality Monitoring Network of the National Institute of Public Health and the Environment (RIVM). Each of these monitoring networks has one site in the center of Rotterdam. Poisson regression models, including weather variables (temperature and relative humidity) and adjustments for long-term trends, seasonal trends, and influenza incidence, were used to examine the association between air pollutants and mortality. The authors note that the association between temperature and mortality in the Netherlands is "highly nonlinear." They therefore use nonparametric smoothers to adjust for temperature. Although concentrations of TSP, black smoke, $SO_2$ and CO were all positively correlated, ozone concentration was negatively correlated with the other pollutants. Ozone, lagged one day, was significantly associated with mortality in a two pollutant model that

included TSP. The relative risk for an increase from the 5th percentile to the 95th percentile ozone concentration (an increase of 34.2 ppb) was 1.06 (95% C.I. = [1.01, 1.11]).

### 3.7 Ito and Thurston, 1996. Daily $PM_{10}$/mortality associations: An investigation of at-risk subpopulations

This study analyzed the relationship between daily mortality and air pollution in Cook County, Illinois, from 1985 to 1990 for the total population and for racial and gender subpopulations. (Results discussed here are for the whole population.) Because Cook County encompasses the city of Chicago, it has the third largest urban population in the nation. Daily death counts were obtained from the National Center for Health Statistics (NCHS). Accidental deaths and deaths occurring outside the county of residence were excluded. Data on $PM_{10}$ (from six sites), $SO_2$ (from five sites), CO (from three sites), and $O_3$ (from five sites) were obtained from EPA's Aerometric Information Retrieval System (AIRS). The authors used Poisson regression models, including weather and pollutant variables, and sine/cosine series to adjust for long-term and seasonal trends, a linear time trend variable, and day-of-week dummy variables. The authors investigated several different specifications of temperature and found that a parabolic, "dual-lag" structure fit the best. This adjustment for temperature was used in all the analyses. Among the pollution variables, $PM_{10}$ and $O_3$ (2-day averages) were most consistently associated with mortality. In single pollutant models, with weather and time- and seasonal-trend adjustments, $PM_{10}$ and $O_3$ were each significantly associated with mortality. When $PM_{10}$ and $O_3$ were both in the same model, the relative risks associated with each were slightly smaller but still statistically significant [for $PM_{10}$, RR = 1.04; 95% C.I. = (1.01, 1.07) for an increase of 100 $\mu g/m^3$; for ozone, RR = 1.07 95% C.I. = (1.01, 1.12) for an increase of 100 ppb]. $PM_{10}$ and $O_3$ were negatively correlated (r = -0.37).

### 3.8 Moolgavkar et al., 1995. Air pollution and daily mortality in Philadelphia

This study analyzed the relationship between daily mortality and air pollution in Philadelphia from 1973 through 1988. Daily death counts were obtained from the National Center for Health Statistics. Accidental deaths and suicides were excluded. In contrast to some other studies, however, these authors chose not to exclude deaths in Philadelphia of nonresidents or deaths of Philadelphia residents that occurred outside of Philadelphia. Air pollution measurements were obtained from EPA's AIRS. Daily averages (averaged over all monitors) of TSP, $SO_2$, and $O_3$ were used in the analyses. All analyses used Poisson regression models, which included quintiles of temperature and indicators for years. Most analyses were done separately by season. In single pollutant models, ozone was associated with mortality only in the summer, defined as June, July, and August [RR = 1.15, 95% C.I. = (1.09, 1.21) for an increase of 100 ppb in the previous day's ozone]. The summertime association of ozone with mortality persisted, however, when the other two pollutants were added to the model [RR = 1.15, 95% C.I. = (1.07, 1.24)]. One analysis treated the entire dataset of 16 years as a single time series (not separated into seasons). In this analysis, indicator variables were used to adjust for seasons and for years, and quintiles of temperature within season were also included. All three pollutants were included

in the model. The relative risk associated with a 100 ppb increase in the previous day's ozone was 1.063 [95% C.I. = (1.018, 1.108)].

### 3.9 Samet et al., 1996, 1997. Particulate air pollution and daily mortality: Analyses of the effects of weather and multiple air pollutants

This study reports the results of Phase I.B of the Particle Epidemiology Evaluation Project, sponsored by the Health Effects Institute (HEI). Analyses were based on time series of mortality and air pollution data from 1974 through 1988 in Philadelphia. Data on TSP, $SO_2$, $NO_2$, CO, and $O_3$ were obtained from EPA's AIRS. The concentrations of TSP, $SO_2$, $NO_2$, and CO were "moderately" correlated with one another; ozone concentration was correlated with concentrations of the other pollutants to a much lesser degree. The signs of the correlations of ozone and the other pollutants varied by season. Ozone was negatively correlated with each of the other pollutants in the winter and positively correlated with each in the summer. (The signs of the correlations for spring and fall varied by pollutant.) Ozone was, in general, not highly correlated with TSP. The correlations of largest magnitude were in the winter and summer (-36.7 in winter, and 36.8 in summer). Based on preliminary explorations, two-day averages (i.e., averages of same day and previous day pollutant levels) were used for all pollutants. Although several models were analyzed, all were of the Poisson regression form.

Particular emphasis was placed on adjustments for weather. In the final model, a nonlinear adjustment for temperature was approximated by four linear terms corresponding to specified temperature cutpoints. Multipollutant models included TSP, $SO_2$, $NO_2$, CO, and $O_3$ and controlled for weather and long-term trends. Unmeasured time trends were controlled for by smoothing spline functions of time. Based on a series of preliminary analyses, the final model included separate effects of TSP for three age groups, and all-age effects of each of the other pollutants, as well as variables for weather and time trends. The relative risk of ozone-related mortality associated with an increase of one interquartile range (20.2 ppb) of ozone reported from this model was 1.019 [95% C.I. = (1.007, 1.032)].

## 4.0 QUANTITATIVE APPROACH

The basic approach of the quantitative analysis is to use the selected ozone-mortality concentration-response relationships and their statistical confidence intervals to estimate a probability distribution of expected national incidence of ozone-related mortality associated with changes in ozone concentrations resulting from specified proposed NAAQS. The analysis may be thought of as having three basic steps. The first step is the selection of study results to include in the analysis, described in Section 2.

In the second step, a distribution of expected national incidence of ozone-related mortality associated with changes in ozone concentrations resulting from a specified proposed NAAQS is derived from each study. Given the ozone coefficient and standard error from a study, and given the appropriate air quality data, a distribution of the national incidence of ozone-related mortality is derived. This distribution describes the probability that the national incidence

of ozone-related mortality associated with a given proposed NAAQS falls within any specified range, if the ozone coefficient appropriate for the entire nation is the ozone coefficient in the study location. If there are N studies in the quantitative analysis, N such distributions of national ozone-related mortality incidence are derived in step 2.

In the third step, a *single* distribution of expected national incidence of ozone-related mortality associated with changes in ozone concentrations resulting from a specified proposed NAAQS is derived from the N distributions derived in step 2. The first step, study and result selection, was described in Section 2. The second and third steps, and the basic assumptions underlying the quantitative analysis, are described below.

### 4.1 Assumption about underlying variability in the relationship between ozone and daily mortality: the fixed effects model versus the random effects model

It is possible that the relationship between ozone and daily mortality is the same everywhere — i.e., that there is a single ozone coefficient that all studies are attempting to estimate. If this is the case, differences in ozone coefficients reported by studies conducted in different locations are due to sampling error and differences in study design. If we believe this model, then we want an estimate of the ozone coefficient and a standard error of that estimate.

It is also possible, however, that the relationship between ambient ozone concentrations and daily mortality differs from one location to another (for example, because of differences in population composition and behavior patterns that may affect susceptibility and exposure to outdoor ozone). If this is the case, differences among reported coefficients may be due not only to sampling error and differences in study design but also to the fact that the studies are estimating different parameters. This model is more plausible and more general than the model of a single ozone coefficient that applies everywhere. (The model of a single ozone coefficient may be thought of as a special case of the general model — the case in which the variability among ozone coefficients is zero.)

The model that assumes that there is a single ozone coefficient in the concentration-response function is called the fixed effects model. The model that allows the possibility that the estimates from different studies may in fact be estimates of different ozone coefficients, rather than just different estimates of a single ozone coefficient, is called the random effects model. The way the results from different studies are combined (in particular, the way different studies are weighted) to obtain a single estimate (of the one ozone coefficient, under the fixed effects model; of the mean of the ozone coefficients, under the random effects model) will depend on the underlying model assumed. This is explained more fully, and an example is given, in the appendix.

A random effects model is the more reasonable model in this situation. Under this model, there is a distribution of ozone coefficients throughout the United States. The mean of this distribution may be used in a national analysis and the distribution itself may be used to

characterize the uncertainty surrounding a "national coefficient."[1] To use the random effects model properly requires that (1) each input study represents a single location, (2) each location is represented only once, and (3) the estimated relationship is between the same variables in all studies — that is, that the ozone-mortality coefficients from the studies are all comparable.

Study selection Criterion 7 ensures that the first condition is met. There are, however, several studies in the set of selected studies that estimated the ozone-mortality relationship in Philadelphia in the same or overlapping time periods. The second condition therefore is not met. To meet the second condition, the Philadelphia results from several studies are pooled, as described in Section 4.2.

Although study selection Criteria 1, 5, and 6 are designed to ensure that the third condition is met, the ozone-mortality coefficients from the selected studies are still not entirely comparable. Some studies estimate the relationship between mortality and daily 1-hour maximum ozone whereas other studies estimate the relationship between mortality and daily (or some other) average ozone. This problem can be solved, however, by replacing the reported ozone coefficient for each study with the national incidence of ozone-related mortality that would be predicted by using that ozone coefficient and the appropriate ozone averaging time for that coefficient. This translates all results into "national incidence space" so that the results from different studies are comparable and can be used to estimate a distribution of national ozone-related mortality incidence. The method for doing this is described in Section 4.3.

## 4.2 Between-study redundancy: avoiding over-representing a single location

Two studies satisfying the study selection criteria listed in Section 2 have estimated a relationship between ozone and daily mortality in Philadelphia for the same or overlapping time periods. Including both of them in the Monte Carlo procedure described below would be giving Philadelphia twice the weight of other locations in that procedure. The two Philadelphia studies and their time periods are:

- Moolgavkar et al., 1995 (1973-1988)
- Samet et al., 1996, 1997 (1974-1988).

To include Philadelphia as one of the locations on which a probability distribution for the national ozone-related mortality incidence is based, the Monte Carlo procedure (as described in

[1]Although each county in the United States may have its own ozone-mortality coefficient, it is infeasible to use county-specific coefficients in a national analysis. Because the national incidence of ozone-related mortality is a continuous function of these ozone-mortality coefficients, it can be shown that there exists a coefficient that, if applied in all counties, would yield the same result as the set of county-specific coefficients (Intermediate Value Theorem). Although this coefficient is unknown, a good candidate for this value is the mean of the distribution of ozone-mortality coefficients.

Section 4.4 below) was carried out first on the two Philadelphia studies to produce a Philadelphia-based probability distribution of national ozone-related mortality incidence. This Philadelphia-based distribution was then used along with the other location-specific distributions generated in the first step of the Monte Carlo procedure described in Section 4.4.

## 4.3 Using studies that measured daily 1-hour maximum ozone and studies that measured daily average ozone (or some variant of the daily average)

Among the 9 studies that satisfy the selection criteria, 4 measured daily 1-hour maximum ozone concentrations and 5 measured daily average (or some other average) ozone concentrations. In order to aggregate the results from these studies, a conversion to one type of measure (i.e., either daily 1-hour maximum or daily average, but not both) is necessary. The peak-to-mean ratios necessary to make such conversions, however, are not available for all of the study locations.

Given that 1-hour maximum and daily average ozone modeling data are both available for the air quality scenarios being evaluated, there is an alternative to converting to either type of ozone measure. Using the ozone data appropriate for a selected study (either 1-hour maximum or daily average) and the ozone coefficient reported by the study, a national ozone-related mortality incidence can be generated. Further, given the standard error of the reported ozone coefficient as well, a distribution of national ozone-related mortality incidences can be generated. This procedure (described more fully in step 1 of Section 4.4 below) converts all study-specific ozone results, some of which correspond to 1-hour maximum ozone levels and some of which correspond to daily average ozone levels, to study-specific national ozone-related mortality incidence results.

## 4.4 The Monte Carlo method of estimating a probability distribution of the national incidence of ozone-related daily mortality

Given a set of studies, some of which use 1-hour maximum ozone and some of which use daily average ozone, the following steps will be used to aggregate the results of these studies to estimate a probability distribution for the national ozone-related incidence of mortality (deaths avoided) corresponding to a given increase (decrease) in ozone concentrations[2]:

[2]Note that all discussion of national ozone-related mortality is specific to a given set of changes in ozone concentrations throughout the United States. A different set of changes in ozone concentrations would result in a different national ozone-related mortality.

1. *For each acceptable study,[3] estimate the probability distribution of national ozone-related mortality incidence that would be predicted, given the ozone coefficient and the reported standard error of that coefficient from the study*. The proposed method to do this is as follows: Let β denote the reported ozone coefficient from the study, and let s.e.(β) denote the reported standard error of the estimate of the ozone coefficient. Then a normal distribution with mean equal to β and standard deviation equal to s.e.(β) describes the probability distribution of what the ozone coefficient is in the location in which the study was conducted. Using the ozone data appropriate to the study (i.e., either daily 1-hour maximum or daily average), calculate the national ozone-related mortality incidence that would be predicted using the (n - 0.5)th percentile of the normal probability distribution described above, for n = 1, 2, 3, ..., 100. That is, calculate the national mortality incidence that would be predicted by using the 0.5th percentile point of the distribution of β's implied by the study, the 1.5th percentile point, and so on.

This step puts all studies, whether they use daily 1-hour maximum ozone or daily average ozone, into "national mortality incidence space" so that they are all comparable. That is, this step produces for each study a probability distribution of national ozone-related mortality incidence corresponding to the probability distribution of ozone coefficients based on the study's estimate of the coefficient and standard error of the estimate.

2. *Generate a single probability distribution of national ozone-related mortality incidence, based on the results of all the acceptable studies*. Such a single distribution is generated from the study-specific distributions by Monte Carlo methods. On each of many iterations, an estimate of the national ozone-related mortality incidence is generated by the following two-step procedure:

> 2a. *Randomly select a study from the set of acceptable studies, using random effects weights*. The probability of selection of a study is a function of both the variance of the estimate from the study (the within-study variance) and the variance among estimates from different studies (between-study variance). This random effects weighting is described in the attached appendix. To calculate random effects weights, both the within-study variance of each study and the between-study variance will have to be calculated in national incidence space — that is, from the distributions of national incidences generated in step 1.
>
> Other considerations could conceivably be incorporated into the weighting scheme (for example, the representativeness of the study location for an analysis of the ozone-mortality relationship within the United States). There is insufficient information, however, to derive nonarbitrary weights that would incorporate such potential considerations. (The degree to which the representativeness of non-U.S. locations is a

[3]Two of the acceptable studies (i.e., studies that satisfy the study selection criteria) were conducted in the same location. This issue of location redundancy is discussed in Section 4.2. The Monte Carlo procedure assumes that each location is represented only once.

concern is itself unclear.) Therefore the random effects weights described in the appendix will be used.

2b. *Randomly select an estimate of the national ozone-related mortality from the distribution of estimates (corresponding to the 100 percentile points) derived in step 1 for the selected study*. The normal probability distribution of incidence for each study is approximated by a histogram in which each bar is centered at one of 100 percentile points (0.5th percentile, 1.5th percentile, 2.5th percentile, ..., 99.5th percentile). Each percentile point therefore has a 1/100 probability of selection — i.e., the random selection in step 2b is from a discrete uniform distribution. (Note that each percentile point is at the center of a segment supporting 1/100th of the probability mass of the normal distribution. Therefore the percentile points are not evenly spaced. They get closer together as they approach the mean of the normal distribution.)

Repeating steps 2a and 2b many times will generate a probability distribution of estimated national ozone-related mortality incidence. The mean of this distribution is the same as the random effects meta-analysis estimate of the mean (see appendix). The shape of the distribution, however, will depend on the information in the underlying studies — how different their estimates are from each other and the relative variances around those estimates. An alternative approach would be to impose a shape on the distribution (e.g., a normal distribution or a beta distribution). The Monte Carlo approach is preferable, however, because it generates a distribution that is most consistent with the existing information.

Probability statements about the national ozone-related mortality incidence can be based on this distribution. For example, if the 5th percentile point of the distribution is denoted as $m_{0.05}$, then there is a 5% probability that the national ozone-related mortality incidence is less than $m_{0.05}$. Similarly, if the 95th percentile point of this distribution is denoted as $m_{0.95}$, then there is a 95% probability that the national ozone-related mortality incidence is less than $m_{0.95}$. There is, then, a 90% probability that the national ozone-related mortality incidence is within the interval $[m_{0.05}, m_{0.95}]$.

### 4.5 Treatment of incidence values less than zero

The resulting (Monte Carlo) distribution of national ozone-related mortality incidence is a composite picture of what the available information tells us about the relationship between a given change in ozone concentrations across the United States and the corresponding change in national mortality. This distribution, however, will have some probability mass to the left of zero, because some of the studies reported statistically insignificant relative risks (or ozone coefficients), and a few studies actually reported (statistically insignificant) relative risks less than 1.0 (or, equivalently, negative ozone coefficients).

It is biologically implausible that exposure to ozone is beneficial. The question, then, is what to do with the probability mass to the left of zero. Redistributing the probability mass below zero to be at zero guarantees that any estimate will be nonzero and that the mean of the

distribution will be positive. This, however, produces a biased estimate of the true mean ozone-related national mortality. (Even if the truth is that the ozone-mortality effect is zero, this procedure guarantees a positive estimate.) The point estimate of national ozone-related mortality, then, should be based on the unadjusted distribution.

For the purpose of making probabilistic statements, however, it is reasonable to redistribute the probability mass to the left of zero to be at zero. Suppose, for example, that 3% of the probability distribution of national ozone-related mortality incidence is to the left of zero. The reasonable inference is that, based on the available information, there is a 3% chance that exposure to ozone has no effect on the risk of premature mortality. Similarly, if 20% of the probability distribution is below zero, then it is reasonable to infer that, based on the available information, there is a 20% chance that exposure to ozone has no effect on the risk of premature mortality. If the 5th percentile of the distribution is to the right of zero, then any probability mass to the left of zero will have no impact on anything that is likely to be reported in the quantitative analysis.

## 5.0 REFERENCES


Anderson, H.R., A. Ponce de Leon, J.M. Bland, J.S. Bowers, and D.P. Strachan. 1996. Air Pollution and Daily Mortality in London: 1987-1992. *British Medical Journal* 312: 665-669.

Cifuentes, L.A., and L. Lave. 1997. Association of Daily Mortality and Air Pollution in Philadelphia, 1983-1988. Submitted to: *Journal of Air Waste Management Association.*

DerSimonian, R., and N. Laird. 1986. Meta-Analysis in Clinical Trials. *Controlled Clinical Trials* 7: 177-188.

Dockery, D.W., J. Schwartz, and J.D. Spengler. 1992. Air Pollution and Daily Mortality: Associations with Particulates and Acid Aerosols. *Environmental Research* 59: 362-373.

Hill, A.B. 1965. The Environment and Disease: Association or Causation? *Proc. R. Soc. Med.* 58: 295-300.

Hoek, G., J.D. Schwartz, B. Groot, and P. Eilers. 1997 (in press). Effects of Ambient Particulate Matter and Ozone on Daily Mortality in Rotterdam, the Netherlands. *Archives of Environmental Health.*

Ito, K., and G.D. Thurston. 1996. Daily $PM_{10}$/Mortality Associations: An Investigation of At-Risk Subpopulations. *Journal of Exposure Analysis and Environmental Epidemiology* 6(1): 79-225.

Katsouyanni, K, A. Pantazopoulou, G. Touloumi, I. Tselepidaki, K. Moustris, D. Asimakopoulos, G. Poulopoulou, and D. Trichopoulos. 1993. Evidence for Interaction Between Air Pollution and High Temperature in the Causation of Excess Mortality. *Archives of Environmental Health* 48(4): 235-242.

Kinney, P.L., and H. Ozkaynak. 1991. Associations of Daily Mortality and Air Pollution in Los Angeles County. *Environmental Research* 54: 99-120.

Kinney, P.L., and H. Ozkaynak. 1992. Associations Between Ozone and Daily Mortality in Los Angeles and New York City. *American Review of Respiratory Disease* 145: A95 (Abstract).

Kinney, P.L., K. Ito, and G.D. Thurston. 1995. A Sensitivity Analysis of Mortality/$PM_{10}$ Associations in Los Angeles. *Inhalation Toxicology* 7: 59-69.

Loomis D.P., V.H. Borja-Aburto, S.I. Bangdiwala, and C.M. Shy. 1996. Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis. Health Effects Institute Research Report Number 75, October 1996.

Moolgavkar, S.H., E.G. Luebeck, T.A. Hall, and E.L. Anderson. 1995. Air Pollution and Daily Mortality in Philadelphia. *Epidemiology* 6(5): 476-484.

Ostro, B.D. 1995. Fine Particulate Air Pollution and Mortality in Two Southern California Counties. *Environmental Research* 70: 98-104.

Ostro B.D., J.M. Sanchez, C. Aranda, and G.S. Eskeland. 1996. Air Pollution and Mortality: Results from a Study of Santiago, Chile. *Journal of Exposure Analysis and Environmental Epidemiology* 6: 97-114.

Ozkaynak, H., J. Xue, P. Severance, R. Burnett, and M. Raizenne. 1995. Associations Between Daily Mortality, Ozone and Particulate Air Pollution in Toronto, Canada. Paper presented at the Colloquium on Particulate Air Pollution, Irvine, CA, January 24-25.

Saldiva, P.H., A.J. Lichtenfels, P.S. Paiva, I.A. Barone, M.A. Martins, E. Massad, J.C. Pereira, V.P. Xavier, J.M. Singer, and G.M Böhm. 1994. Association Between Air Pollution and Mortality due to Respiratory Diseases in Children in Sao Paulo, Brazil: A Preliminary Report. *Environmental Research* 65: 218-225.

Saldiva, P.H., C.A. Pope, 3rd, J. Schwartz, D.W. Dockery, A.J. Lictenfels, J.M. Salge, I. Barone, and G.M. Barone. 1995. Air Pollution and Mortality in Elderly People: A Time-Series Study in Sao Paulo, Brazil. *Archives of Environmental Health* 50(2): 159-163.

Samet, J.M., S.L. Zeger, J.E. Kelsall, and J. Xu. 1996. Air Pollution and Mortality in Philadelphia, 1974-1988. Report to the Health Effects Institute on Phase IB: Particle Epidemiology Evaluation Project, March 25, 1996 (draft, accepted for publication).

Samet, J.M., S.L. Zeger, J.E. Kelsall, J. Xu, and L.S. Kalkstein. 1997. Particulate Air Pollution and Daily Mortality: Analysis of the Effects of Weather and Multiple Air Pollutants. The Phase I.B Report of the Particle Epidemiology Evaluation Project. Health Effects Institute, March 1997.

Sartor, F., R. Snacken, C. Demuth, and D. Walckiers. 1995. Temperature, Ambient Ozone Levels, and Mortality during Summer 1994, in Belgium. *Environmental Research* 70: 105-113.

Schwartz, J. 1991. Particulate Air Pollution and Daily Mortality in Detroit. *Environmental Research* 56: 204-213.

Shumway, R.H., A.S. Azari, and Y. Paurtan. 1988. Modeling Mortality Fluctuations in Los Angeles as Functions of Pollution and Weather Effects. *Environmental Research* 56: 204-213.

Simpson, R.W., G. Williams, A. Petroeschevsky, G. Morgan, and S. Rutherford. 1997. The Association Between Outdoor Air Pollution and Daily Mortality in Brisbane, Australia. Submitted to: *Archives of Environmental Health*.

Sunyer, J., J. Castellsagué, M. Saez, A. Tobias, and J.M. Anto. 1996. Air Pollution and Mortality in Barcelona. *Journal of Epidemiology and Community Health* 50 (Suppl. 1): S76-S80.

Thurston, G. 1997. Ozone Air Pollution and Human Mortality. Paper to be presented at the Air and Waste Management Association's 90th Annual Meeting and Exhibition, June 8-13, 1997, Toronto, Ontario, Canada. EPA Docket No. A-95-58 IV-D-2228.

Touloumi et al. 1997 (in press). Short Term Effects of Ambient Oxidants Exposure on Mortality: A Combined Analysis Within the APHEA Project. *American Journal of Epidemiology*.

U.S. Environmental Protection Agency. 1996. *Air Quality Criteria for Ozone and Related Photochemical Oxidants.* Vol. I-III. EPA/600/P-93/004a,b,cF. Office of Research and Development, Washington, DC, July.

U.S. Environmental Protection Agency. 1996. *Air Quality Criteria for Particulate Matter.* Vol. I-III. EPA/600/P-95/001a,b,cF. Office of Research and Development, Washington, DC, April.

Verhoeff, A.P., G. Hoek, J. Schwartz, and J.H. van Wijnen. 1996. Air Pollution and Daily Mortality in Amsterdam. *Epidemiology* 7(3): 225-230.

Wyzga, R.E., and F.W. Lipfert. 1995. Temperature-Pollution Interactions with Daily Mortality in Philadelphia. ***In:*** Particulate Matter: Health and Regulatory Issues: Proceedings of an International Specialty Conference Sponsored by the Air and Waste Management Association; April; Pittsburgh, PA., pp. 3-42.

Wyzga, R.E., and F.W. Lipfert. 1996. Ozone and Daily Mortality: The Ramifications of Uncertainties and Interactions and Some Initial Regression Results. ***In:*** Vostal, J.J., ed., Tropospheric Ozone: Critical Issues in the Regulatory Process: Proceedings of a Specialty Conference Sponsored by the Air and Waste Management Association; May 1994; Orlando, FL., pp. 453-487.

Zmirou, Z.D., T. Barumandzadeh, F. Balducci, P. Ritter, G. Laham, and J.P. Ghilardi. 1996. Short Term Effects of Air Pollution on Mortality in the City of Lyon, France, 1985-1990. *Journal of Epidemiology and Community Health* 50 (Suppl. 1): S30-S35.

## Appendix J.2: Pooling the Results of Different Studies

### J.2.0 Introduction

Many studies have attempted to determine the influence of ozone pollution on human health. Usually this involves estimation of a parameter β in a concentration-response function, which may be linear or nonlinear, as discussed above. Each study provides an estimate of β, along with a measure of the uncertainty of the estimate. Because uncertainty decreases as sample size increases, combining data sets is expected to yield more reliable estimates of β. Combining data from several comparable studies in order to analyze them together is often referred to as meta-analysis.

For a number of reasons, including data confidentiality, it is often impractical or impossible to combine the original data sets. Combining the *results* of studies in order to produce better estimates of β provides a second-best but still valuable way to synthesize information (DerSimonian and Laird, 1986). This is referred to as "pooling results" in this report. Pooling requires that all of the studies contributing estimates of β use the same functional form for the concentration-response function. That is, the β's must be measuring the same thing.

One method of pooling study results is simply averaging all reported β's. This has the advantage of simplicity, but the disadvantage of not taking into account the uncertainty of each of the estimates. Estimates with great uncertainty are given the same weight as estimates with very little uncertainty. For example, consider the three studies whose results are presented in Table J.1.

**Table J.1 Three Sample Studies**

| Study | Estimate of β | Standard Deviation | Variance |
|---|---|---|---|
| Study 1 | 0.75 | 0.35 | 0.1225 |
| Study 2 | 1.25 | 0.05 | 0.0025 |
| Study 3 | 1.00 | 0.10 | 0.0100 |

The average of the three estimates is 1.0. However, the Study 2 estimate has much less uncertainty associated with it (variance = 0.0025) than either the Study 1 or Study 3 estimates. It seems reasonable that a pooled estimate that combines the estimates from all three studies should therefore give more weight to the estimate from the second study than to the estimates from the first and third studies. A common method for weighting estimates involves using their variances. Variance takes into account both the consistency of data and the sample size used to obtain the estimate, two key factors that influence the reliability of results.

The exact way in which variances are used to weight the estimates from different studies in a pooled estimate depends on the underlying model assumed. The next section discusses the two basic models that might underlie a pooling and the weighting scheme derived from each.

### J.2.1 The fixed effects model

The fixed effects model assumes that there is a single true concentration-response relationship and therefore a single true value for the parameter β. Differences among β's reported by different studies are therefore simply the result of sampling error. That is, each reported β is an estimate of the same underlying parameter. The certainty of an estimate is reflected in its variance (the larger the variance, the less certain the estimate). Pooling that assumes a fixed effects model therefore weights each estimate under consideration in proportion to the inverse of its variance.

Suppose there are n studies, with the ith study providing an estimate $\beta_i$ with variance $v_i$ (I = 1, ..., n). Let

$$S = \Sigma \frac{1}{v_i},$$

denote the sum of the inverse variances. Then the weight, $w_i$, given to the ith estimate, $\beta_i$, is

$$w_i = \frac{1/v_i}{S} .$$

This means that estimates with small variances (i.e., estimates with relatively little uncertainty surrounding them) receive large weights, and those with large variances receive small weights.

The estimate produced by pooling based on a fixed effects model is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{fe} = \Sigma\ w_i * \beta_i .$$

The variance associated with this pooled estimate is the inverse of the sum of the inverse variances:

$$v_{fe} = \frac{1}{\Sigma\ 1/v_i} .$$

Table J.2 shows the relevant calculations for this pooling for the three sample studies summarized in Table J.1.

**Table J.2 Fixed Effect Model Calculations**

| Study | $\beta_i$ | $v_i$ | $1/v_i$ | $w_i$ | $w_i*\beta_i$ |
|---|---|---|---|---|---|
| 1 | 0.75 | 0.1225 | 8.16 | 0.016 | 0.012 |
| 2 | 1.25 | 0.0025 | 400 | 0.787 | 0.984 |
| 3 | 1.00 | 0.0100 | 100 | 0.197 | 0.197 |
| Sum | | | $\sum$ = 508.16 | $\sum$ = 1.000 | $\sum$ = 1.193 |

The sum of weighted contributions in the last column is the pooled estimate of β based on the fixed effects model. This estimate (1.193) is considerably closer to the estimate from Study 2 (1.25) than is the estimate (1.0) that simply averages the study estimates. This reflects the fact that the estimate from Study 2 has a much smaller variance than the estimates from the other two studies and is therefore more heavily weighted in the pooling.

The variance of the pooled estimate, $v_{fe}$, is the inverse of the sum of the inverse variances, or 0.00197. (The sums of the $\beta_i$ and $v_i$ are not shown, since they are of no importance. The sum of the $1/v_i$ is S, used to calculate the weights. The sum of the weights, $w_i$, I = 1, ..., n, is 1.0, as expected.)

### J.2.2 The random effects model

An alternative to the fixed effects model is the random effects model, which allows the possibility that the estimates $\beta_i$ from the different studies may in fact be estimates of different parameters, rather than just different estimates of a single underlying parameter. In studies of the effects of ozone on mortality, for example, if the behavior or susceptibility of populations varies among study locations, the underlying relationship between mortality and ambient ozone concentrations may be different from one study location to another. (Suppose, for example, people in one location spend substantially more time outdoors than people in another location; this would violate the assumption of the fixed effects model.)

The following procedure can test whether it is appropriate to base the pooling on the random effects model (versus the fixed effects model):

A test statistic, $Q_w$, the weighted sum of squared differences of the separate study estimates from the pooled estimate based on the fixed effects model, is calculated as:

$$Q_w = \Sigma_i \frac{1}{v_i} (\beta_{fe} - \beta_i)^2.$$

Under the null hypothesis that there is a single underlying parameter, β, of which all the $\beta_i$s are estimates, $Q_w$ has a chi-squared distribution with n-1 degrees of freedom. (Recall that n is the

number of studies in the meta-analysis.) If $Q_w$ is greater than the critical value corresponding to the desired confidence level, the null hypothesis is rejected. That is, in this case the evidence does not support the fixed effects model, and the random effects model is assumed, allowing the possibility that each study is estimating a different β.

The weights used in a pooling based on the random effects model must take into account not only the within-study variances (used in a meta-analysis based on the fixed effects model) but the between-study variance as well. These weights are calculated as follows:

Using $Q_w$, the between-study variance, $\eta^2$, is:

$$\eta^2 = \frac{Q_w - (n-1)}{\Sigma 1/v_i - \frac{\Sigma 1/v_i^2}{\Sigma 1/v_i}} \quad .$$

It can be shown that the denominator is always positive. Therefore, if the numerator is negative (i.e., if $Q_w < n-1$), then $\eta^2$ is a negative number, and it is not possible to calculate a random effects estimate. In this case, however, the small value of $Q_w$ would presumably have led to accepting the null hypothesis described above, and the meta-analysis would be based on the fixed effects model. The remaining discussion therefore assumes that $\eta^2$ is positive.

Given a value for $\eta^2$, the random effects estimate is calculated in almost the same way as the fixed effects estimate. However, the weights now incorporate both the within-study variance ($v_i$) and the between-study variance ( $\eta^2$). Whereas the weights implied by the fixed effects model used only $v_i$, the within-study variance, the weights implied by the random effects model use $v_i + \eta^2$.

Let $v_i^* = v_i + \eta^2$. Then

$$S^* = \Sigma \frac{1}{v_i^*} \ ,$$

and

$$w_i^* = \frac{1/v_i^*}{S^*} \quad .$$

The estimate produced by pooling based on the random effects model, then, is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{rand} = \Sigma\ w_i^* * \beta_i .$$

The variance associated with this random effects pooled estimate is, as it was for the fixed effects pooled estimate, the inverse of the sum of the inverse variances:

$$v_{rand} = \frac{1}{\Sigma\ 1/v_i^*} .$$

The weighting scheme used in a pooling based on the random effects model is basically the same as that used if a fixed effects model is assumed, but the variances used in the calculations are different. This is because a fixed effects model assumes that the variability among the estimates from different studies is due only to sampling error (i.e., each study is thought of as representing just another sample from the same underlying population), whereas the random effects model assumes that there is not only sampling error associated with each study, but that there is also between-study variability — each study is estimating a different underlying β. Therefore, the sum of the within-study variance and the between-study variance yields an overall variance estimate.

### J.2.3 An example

This section demonstrates the relevant calculations for pooling using the example in Table J.1 above.

First calculate $Q_w$, as shown in Table J.3.

**Table J.3 Calculation of $Q_w$**

| Study | $\beta_i$ | $1/v_i$ | $1/v_i * (\beta_i - \beta_{fe})^2$ |
|---|---|---|---|
| 1 | 0.75 | 8.16 | 1.601 |
| 2 | 1.25 | 400 | 1.300 |
| 3 | 1.00 | 100 | 3.725 |
| | | | $\sum = Q_w = 6.626$ |

In this example the test statistic $Q_w = 6.626$. The example considers three studies, so $Q_w$ is distributed as a chi-square on two degrees of freedom. The critical value for the 5% level (i.e., corresponding to a 95% level of confidence) for a chi-square random variable on 2 degrees of freedom is 5.99. Because $Q_w = 6.626 > 5.99$, hence the null hypothesis is rejected. That is, the

evidence does not support the fixed effects model. Therefore assume the random effects model is appropriate.

Then calculate the between-study variance:

$$\eta^2 = \frac{6.626 - (3 - 1)}{508.16 - \frac{170066.65}{508.16}} = 0.0267 \quad .$$

From this and the within-study variances, calculate the pooled estimate based on the random effects model, as shown in Table J.4.

**Table J.4 Random Effects Model Calculations**

| Study | $\beta_i$ | $v_i + \eta^2$ | $1/(v_i + \eta^2)$ | $w_i^*$ | $w_i^* \times \beta_i$ |
|---|---|---|---|---|---|
| 1 | 0.75 | 0.1492 | 6.70 | 0.098 | 0.0735 |
| 2 | 1.25 | 0.0292 | 34.25 | 0.502 | 0.6275 |
| 3 | 1.00 | 0.0367 | 27.25 | 0.400 | 0.400 |
| Sum | | | $\sum = 68.20$ | $\sum = 1.000$ | $\sum = 1.101$ |

The random effects pooled estimate, $\beta_{rand}$, is 1.101. It's variance, $v_{rand}$, is 1/(68.2) = 0.015.